

COMMENT

HISTORY The shocking use of soldiers to test chemical weapons p.154



NEUROSCIENCE A corrective primer on the power and limitations of fMRI p.156

CONSERVATION Crackdown on Madagascar's illegal trade in live lemurs p.157

ENERGY Natural gas could help Germany transition to electric cars p.157

SEBASTIEN NOGIER/EPA



Citizens cast votes in a second-round primary of the 2017 French presidential election.

Compare voting systems to improve them

Research is needed on how groups make choices in real situations, write **Guruprasad Madhavan, Charles Phelps and Rino Rappuoli**.

For the second time since 2000 and at least the fourth in the history of the United States, the winner of the presidential election did not win the popular vote. The framers of the US Constitution established the electoral-college system deliberately. To prevent a few large states from dominating outcomes, they blunted the voting power of people within them.

The United Kingdom has a similar system. In 2005, for example, Britain's Labour Party won 55% of the seats in the House of Commons with only 35% of the votes cast.

In the French presidential race, by contrast, the winner always receives a majority of the votes cast. If neither of the two top vote recipients from a first-round election receives a majority, voters choose between the two candidates in a final 'run-off' election. But such methods can lead to a winner who may not be the most preferred candidate overall, depending on how voters pick candidates in the preliminary round.

Which is the best voting system? It depends on the circumstances. The study of choices by groups — social-choice theory

— seeks to understand how different rules affect outcomes, and how well those rules perform against dozens of criteria. But there have been few practical and systematic studies — especially head-to-head comparisons of different rules and how they influence collective decisions, from governments to scientific societies.

POOR UNDERSTANDING

Voting, weighting, ranking, scoring and grading mechanisms are poorly understood by most people who apply them. The ▶

PRIORITIZATION PILOT

One tool, two groups, three vaccines, different choices

At the end of a four-week vaccinology course at Institut Pasteur in Paris in 2016, we asked the 26 participants to rank three hypothetical rotavirus vaccine candidates for development using our software tool, SMART Vaccines. The participants included research scientists, clinicians, doctoral students and programme directors and managers in governments and corporations. We divided them into two groups: one to simulate a health ministry; the other a finance ministry.

We asked each group to select and rank the five attributes most relevant to them, such as cost effectiveness or quality life years gained, from a longer list. The health group shortlisted seven and converged on five through discussion. The finance group winnowed 14 down to five by voting. Both groups ranked their five in importance using two methods: the Borda count, which sums points awarded by each member⁸; and majority judgement, which uses the

median of categorical grades — such as A, B, C, D, E, F — awarded to the options⁴. These ranks were fed into software to calculate scores between zero and 100 — zero representing an ineffective vaccine candidate and a 100 representing a perfect one.

Both groups agreed on the top priority for the vaccine candidates (cases prevented) but differed on which other attributes of the medicines were most important to them. Each group scored the candidates differently, meaning that they would prioritize different ones for development.

Asked which method they preferred, all participants said that majority judgement was easier: it was simpler to assign letter grades than numbers and it tolerated ties. Ranking says nothing about the quality of the choices. For example, four candidates might be graded A, A, D, F and yet ranked 1, 2, 3, 4.

Group	Attributes selected as most important	Priority (Borda ranking)	Priority (majority judgment grading)
Simulated health ministry	Cases prevented per year	1	1
	Quality-adjusted life-years gained	2	3
	Cost-effectiveness	3	2
	Fits into existing immunization schedules	4	4
	Benefits socioeconomically disadvantaged	5	5
Simulated finance ministry	Cases prevented per year	1	1
	Cost-effectiveness	2	2
	Net direct savings of vaccine use per year (in millions)	3	3
	Availability of alternative public-health measures	4	5
	Benefits socio-economically disadvantaged	5	4

► biases and limitations of each method are underappreciated. We pick a good wine with the help of our social network, sommeliers, magazine reviews, point scales and star ratings. But each person weighs attributes differently, leading to different scores for the same wines. Similar issues arise at every level of society, from choosing houses to picking spouses. Science, too, faces collective-choice challenges. Which country should host the proposed International Linear Collider, for instance? Which space missions should NASA undertake? Which medications, foods and consumer goods should be approved?

Bad group decisions harm research, the economy and society. For example, the way in which honorific and professional societies shortlist candidates for elections preserves the current membership structure rather

than diversifying it. Voting rules can aid or discourage public participation — the cornerstone of a democratic process. Some voting methods favour centrists, others extremes.

Improving collective decision-making requires a new type of research: direct comparisons of various voting rules in different settings. Some of this can be done using computer simulations. Much will be learned by testing how decisions are made in real situations.

PICKING WINNERS

Even in this age of machine learning and social media we rely on a handful of basic choice mechanisms. A group may want to select one winner, such as for Pope, or pick the top few of many candidates, for say, a city council or a corporate board. Students are graded to evaluate excellence; apps, hotels,

films and cars are rated by users to provide information to consumers.

All mechanisms have two key features: what type of voter input they use (say, affirmation, or letter or numeric grades) and how those inputs are aggregated (mostly totals, averages or medians). Some methods are more expressive than others. With, say, five candidates, first-past-the-post voters have only six different options or ‘words’ to use, including a blank ballot. A ‘rank order list’ (wherein voters rank options in a numerical order of preference) has 120 different arrangements for the same five candidates. A system with six grades (such as the standard academic A, B, C, D, E, F) provides more than 7,776 combinations.

In the latest US election, the inability of voters to choose between two seemingly unpopular presidential candidates — Hillary Clinton and Donald Trump — may have lowered turnout to less than 60% of eligible voters, and less than in 2008 or 2004 (other factors may have also contributed to the reduced vote totals). Neither of the two major candidates received 17 million votes in the primary elections out of a total of 232 million eligible voters. Thus, only about 7% of eligible voters determined the final candidate slate.

WEIGHING UP METHODS

Social-choice theorists evaluate selection methods using dozens of criteria. For example, ‘join consistency’ requires that an election result is unchanged if two or more districts are merged. As the differing winner of the popular vote in the 2016 US election shows, the US presidential-election system fails this test, as do many other voting systems.

The ‘rank reversal rule’ states that the ranking of any two candidates, A and B, should not change if C enters or leaves the race¹. A notable violation is from 2000, when Green Party candidate Ralph Nader captured a few per cent of the vote in Florida, giving the election to George W. Bush (over Al Gore). Gore would have won either if Nader was not in the race or if Florida had used a run-off election. But a run-off approach can also lead to rank reversal.

Some of these criteria conflict — meeting one guarantees failing another or they confound the electorate’s intention. Many situations invite strategic voting, wherein blocs of voters can alter the outcome by voting for someone other than their true preference. In a run-off ballotage, voters preferring one candidate may elevate a weaker opponent into the final round, assuring the victory of their favourite.

Voting systems that sum or average the points assigned to candidates — such as the rank-order ballots used in Australian elections — exacerbate these risks. Major Australian political parties hand out strategies so

voters may game such opportunities.

Yet there have been few attempts to compare different methods for reaching a group decision. Voting methods can be tested in real elections by getting voters to fill out two types of ballot — an official one using established voting rules and a test ballot with alternative ways for voters to express themselves. One can then ask the voters which method they found easiest to use, which they understood, trusted and possibly preferred to use in future.

Such comparisons in wider political contexts are rare. In one example, members of the Society for Social Choice and Welfare were asked to rank three candidates for their president as well as voting for one using the official voting rule. Later analysis revealed that the different voting rules elected different candidates^{2,3}.

In a larger experiment, voters from several districts in the 2007 French presidential election were asked by researchers Michel Balinski and Rida Laraki to fill out a separate ballot in which they graded each candidate. The researchers tested the time needed to vote (about one minute per ballot), preferences (voters found this ‘majority judgement’ system easy to use and appreciated the increased vocabulary), and how the official vote, majority judgement and other outcomes compared⁴.

The real French presidential election results showed two finalists, who ranked second and third in the polls, in a run-off against each other. The test system would have selected the candidates most preferred in the polls. The difference was not a defect of polling accuracy but rather the result of widespread strategic voting in the first round that was widely discussed before election day in the French media⁴.

We — an engineer, an economist and a biologist — have confronted similar issues in our own work. We developed a decision-support tool for public officials, corporate executives and donors to prioritize potential vaccines for research-and-development investment^{5–7}.

Our software assigns a weight to each of a list of vaccine candidates based on a range of factors, from cost-effectiveness to target population and disease characteristics. The final scores are based on the attributes that a decision-maker values most in domains such as health, economic, demographic, business, operations and policy. And, like most such aids, it is written for individuals, not groups. Almost all its users have asked us for guidance on how to create prioritized lists by combining individual preferences in, say, an advisory committee.

To learn more about how groups use our software, we carried out a trial (see ‘Prioritization pilot’). We compared some common voting methods to see whether the resulting recommendations differed (they

ALBIN LOHR-JONES/PACIFIC PRESS/GETTY



The United Nations General Assembly: collective decision-making in action.

did) and which approach people preferred. Our participants found it easier and more expressive of their views to evaluate the merits of the options using letter grades rather than numerical ranks. Although this is but one example, and not definitive, more such experiments would help to guide groups in other scenarios.

FIELD TRIALS

We cannot help groups and society to choose better selection mechanisms by continuing to study which voting systems pass or fail theoretical criteria. Social-choice research needs to include human factors such as ease

“Voting methods can be tested in real elections by getting voters to fill out two types of ballot .”

and preference of use, and effects on voter participation — a key element in any democratic society. First, it should focus on practical ways to assess the performance and impact of various mechanisms under

different conditions.

Field-testing different selection methods in parallel with real voting situations should determine, for example, how often they produce different results and how much voters understand each approach, as well as what the potential consequences are. Such information can help groups to pick which rule would be best for them. Computer simulations could be used to analyse distributions of public preferences, and also situations in which conflicting voting problems actually occur.

Two good places to start in science, engineering and public health are: peer-review panels that prioritize research proposals for funding decisions; and professional societies choosing members and officers. Testing

in other organizations — from corporate boards to religious congregations and performance awards in sports, literature, arts, and films — would broaden understanding of the human factors at play.

We urge government agencies and philanthropic organizations to fund such practical research. Learning more about choice mechanisms will have widespread social and economic benefits, from better presidents to better science. ■

Guruprasad Madhavan is a senior programme officer in health and medicine at the US National Academies of Sciences, Engineering, and Medicine, Washington, DC, USA. **Charles Phelps** is university professor and provost emeritus at the University of Rochester, New York, USA.

Rino Rappuoli is chief scientist and head of external research and development at GlaxoSmithKline Vaccines, Siena, Italy. e-mail: gmadhavan@nas.edu

1. Arrow, K. J. *Polit. Econ.* **58**, 328–346 (1950).
2. Brahm, S. J. & Fishburn, P. *Soc. Choice Welf.* **18**, 409–414 (2001).
3. Saari, D. G. *Soc. Choice Welf.* **18**, 415–430 (2001).
4. Balinski, M. & Laraki, R. *Majority Judgment: Measuring, Ranking, and Electing* (MIT Press, 2010).
5. Madhavan, G. et al. (eds) *Ranking Vaccines: A Prioritization Framework* (Institute of Medicine, National Academies Press, 2012).
6. Madhavan, G. et al. (eds) *Ranking Vaccines: A Prioritization Software Tool* (Institute of Medicine, National Academies Press, 2013).
7. Madhavan, G. et al. (eds) *Ranking Vaccines: Applications of a Prioritization Software Tool* (Institute of Medicine and National Academy of Engineering, National Academies Press, 2015).
8. de Borda, J.-C. ‘*Mémoire sur les élections au scrutin*’ (*Histoire de l’Academie royal des sciences*, Paris, 1781).

The views expressed here do not necessarily reflect those of the US National Academies of Sciences, Engineering, and Medicine.